JOURNAL OF THE EXPERIMENTAL ANALYSIS OF BEHAVIOR 2004, **82,** 57–69 NUMBER 1 (JULY)

SIGNAL-DETECTION ANALYSES OF CONDITIONAL DISCRIMINATION AND DELAYED MATCHING-TO-SAMPLE PERFORMANCE

BRENT ALSOP

UNIVERSITY OF OTAGO

Quantitative analyses of stimulus control and reinforcer control in conditional discriminations and delayed matching-to-sample procedures often encounter a problem; it is not clear how to analyze data when subjects have not made errors. The present article examines two common methods for overcoming this problem. Monte Carlo simulations of performance demonstrated that both methods introduced systematic deviations into the results, and that there were genuine risks of drawing misleading conclusions concerning behavioral models of signal detection and animal short-term memory.

Key words: simulations, signal detection, delayed matching to sample, conditional discrimination

---

Signal-detection analyses are often used to measure stimulus control and reinforcer control in conditional discriminations and in delayed matching-to-sample (DMTS) procedures. In such procedures, subjects are usually presented with one of two sample stimuli, and they must identify which sample was presented by choosing between two alternative responses. Correct identifications are reinforced on some schedule of reinforcement. It is not uncommon, however, for subjects to make few or no errors of one type or another in such tasks, particularly if the discriminability between the sample stimuli is high or the payoff matrix has produced substantial response bias. Zero responses in error cells of a detection matrix are problematic because they complicate analyses using signal-detection measures of accuracy or response bias (e.g., Alsop & Davison, 1991; Davison & Tustin, 1978; Green & Swets, 1966); for example, the analysis might entail division by zero.

A number of "correction rules" (Hautus, 1995) allow calculation of measures of discriminability and response bias when detection matrices contain zero responses in cells. One correction method, known as the 1/(2N) rule (Macmillan & Kaplan, 1985), replaces extreme proportions of zero or one with values of 1/(2N) and 1-1/(2N), respectively, where N equals the number of trials on which the proportion is based. This method appears to have had little or no use in behavioral studies of signal detection or DMTS performance. The second method simply replaces instances of zero responses with one response. Here, this will be called the 1s-for-0s rule. A number of articles in *JEAB* have used this method (e.g., Alsop & Davison, 1991; Jones & White, 1992; Watson & Blampied, 1989). More recently, some articles in *JEAB* (Davison & Nevin, 1999; Godfrey & Davison, 1998, 1999; Nevin, Milo, Odum, & Shahan, 2003) have advocated, following from Hautus's analysis, a different method to minimize the problem of zero (or very low) error rates. This method, the log-linear rule, involves adding 0.5 to every cell of the detection matrix, and if it is used, then this correction rule should be applied to all data in the analysis, even those cases in which there is no problem of zero errors.

Here I argue that correction methods, including the log-linear rule, need to be treated with more caution. Although they allow *ordinal* comparisons among conditions, they also can introduce systematic deviations into the results, and higher-order quantitative analyses and modeling can produce misleading results.

To assess the effect of applying the correction rules to data from detection procedures, I generated hypothetical data using the Davison and Tustin (1978) behavioral model of signal detection. This model is based on the generalized matching law (e.g., Baum, 1974). Following a presentation of one sample stimulus (S_1), the model predicts that performance can be described by the relation

Thanks to Jeff Miller for his helpful comments concerning the original manuscript, and J.A. Nevin and Celia Lie for later comments. Address correspondence to Brent Alsop, Department of Psychology, University of Otago, Box 56, Dunedin, New Zealand (e-mail: balsop@psy.otago.ac.nz).

$$\frac{B_{11}}{B_{12}} = cd\left(\frac{R_{11}}{R_{22}}\right)^a, \quad (1)$$

and following a presentation of the other sample stimulus (S_2), by the relation

$$\frac{B_{21}}{B_{22}} = c\frac{1}{d}\left(\frac{R_{11}}{R_{22}}\right)^a, \quad (2)$$

where B_{11} and B_{22} are numbers of correct responses and B_{12} and B_{21} are numbers of incorrect responses. The parameter c measures any inherent bias for one of the response alternatives, and the parameter d measures the discriminability between the sample stimuli. The parameter a measures the sensitivity of behavior to changes in the ratio of reinforcers (R_{11}/R_{22}) obtained for correct B_{11} and B_{22} responses. For the generation of the hypothetical data, no inherent bias was assumed (i.e., $c = 1$), and the sensitivity of behavior to changes in the reinforcer ratio, a, was always set to 0.9 (e.g., Davison & McCarthy, 1988). A value of d was selected (e.g., $d = 10$) with a particular ratio of reinforcers (e.g., $R_{11}/R_{22} = 1{:}9$). This allowed calculation of the expected ratio of responses (B_{11}/B_{12} and B_{21}/B_{22}) following S_1 and S_2 presentations using Equations 1 and 2, and these ratios were then converted to proportions. In order to approximate actual performance, Monte Carlo simulations used these expected proportions to generate values of B_{11}, B_{12}, B_{21}, and B_{22} for a predetermined number of S_1 and S_2 trials. The log-linear rule and the 1s-for-0s rule were applied to these data.

Discriminability in Detection and Delayed Matching-to-Sample Procedures

In the first analysis, simulated B_{11}, B_{12}, B_{21}, and B_{22} response totals were generated for five different reinforcer ratios (9:1, 3:1, 1:1, 1:3, 1:9) at each of three different levels of stimulus discriminability (i.e., $d = 10$, 100, and 1000). Initially, the simulations assumed 500 S_1 trials and 500 S_2 trials in each condition. This mimics studies in this area that collect the data from a relatively large number of trials (e.g., Alsop & Davison, 1991; McCarthy & Davison, 1979; McCarthy & Davison, 1980). The two correction rules were then applied to these data. At each level of discriminability, point estimates of discriminability were calculated for each reinforcer ratio at each level of discriminability using the equation

$$\frac{1}{2}\log\left(\frac{B_{11}B_{22}}{B_{12}B_{21}}\right) = \log d, \quad (3)$$

that can be obtained by algebraic combination of Equations 1 and 2 (Davison & Tustin, 1978). Note that Equation 3 predicts that stimulus discriminability should be independent of changes in the reinforcer ratio. These simulations and calculations were repeated 20 times. This entire analysis was then repeated using 80 trials of each stimulus for each condition to mimic studies that have run fewer trials (e.g., Jones & White, 1992; Nevin et al., 2003; Sargisson & White, 2003).

Figure 1 plots the mean estimates of log d across the 20 simulations as a function of the logarithm of the reinforcer ratio. The left panel shows the results using the log-linear rule, and the right panel shows the results using the 1s-for-0s rule. The dashed lines show the true measures of log d, which must be constant across changes in the reinforcer ratio by definition (Equation 3). The filled circles show the mean estimates of log d following 500 trials with each stimulus (1000 trials), and the open circles show the mean estimates following 80 trials with each stimulus (160 trials). Figure 1 shows that both rules increasingly underestimated the true value of log d as the arranged stimulus discriminability increased, and as the reinforcer ratio became more extreme. The number of trials in the simulation also had an effect; fewer trials produced larger underestimations, markedly so in some cases. Overall, the 1s-for-0s rule produced greater underestimations of the true log d than the log-linear rule.

The systematic deviations from the true value of log d shown in Figure 1, especially as the reinforcer ratio was varied, are a problem for empirical analyses of behavioral models of signal detection. Although correction rules allow analysis of the data from conditions when discriminability between the samples is high, they can also introduce quite misleading interactions between parameters; any higher-order quantitative analysis could be compromised. In the simulations shown in Figure 1, for example, systematic changes in log d as a function of the reinforcer ratio

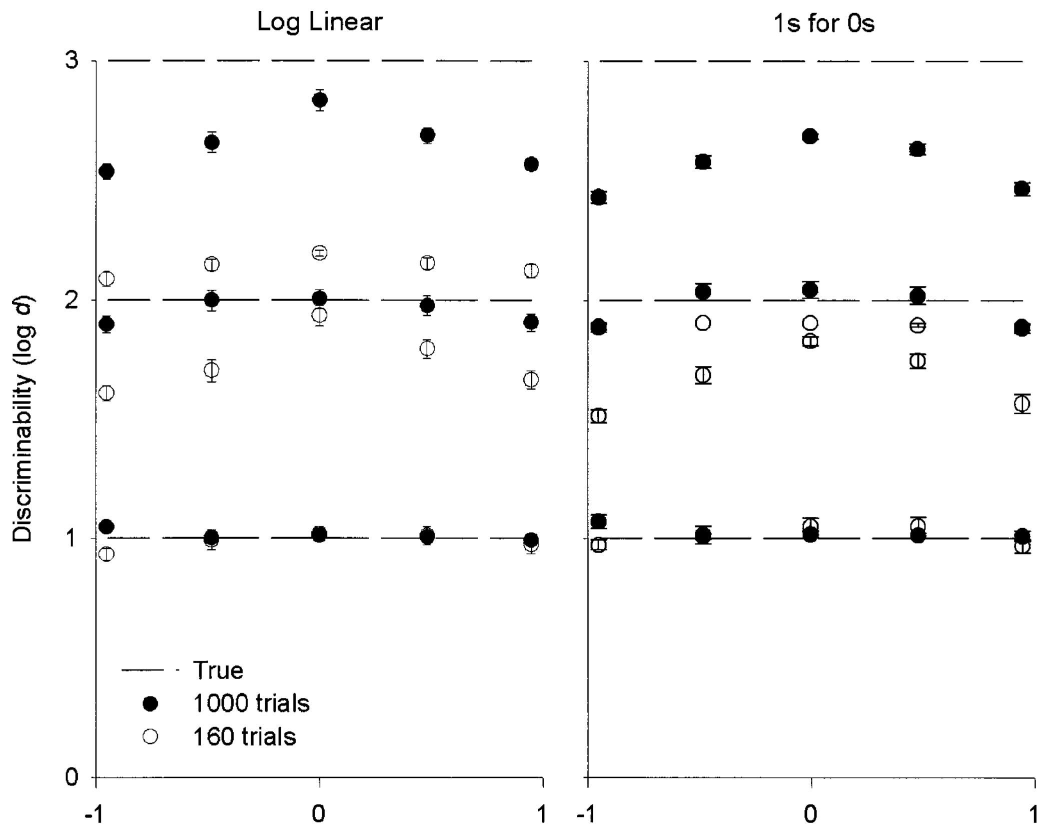

Fig. 1. Mean discriminability between the sample stimuli (and standard errors) across 20 simulations is plotted as a function of the log reinforcer ratio after the log-linear rule (left panel) and the 1s-for-0s rule (right panel) were applied to the simulated data. The dashed line shows true discriminability (see text for details).

would indicate that the Davison and Tustin (1978) model was flawed (e.g., Equation 3) and in this case, that the data would be better described by a model such as that of Alsop and Davison (1991)[1] (e.g., see Figure 6 of Davison & Nevin, 1999).

The underestimation of log d at higher levels of discriminability shown in Figure 1 also has implications for signal-detection analyses of forgetting functions from DMTS procedures. In DMTS procedures, there is a delay between the offset of the sample stimuli and the opportunity to make the choice responses. It is common to arrange a number of these retention intervals (RIs) across trials (e.g., 0 s, 2 s, 4 s, 8 s, and 16 s) in a quasi-random order within each experimental condition. Short-term memory can be studied by comparing the accuracy of performance at the different RIs. Accuracy typically decreases as the RI increases. The leftmost panel of Figure 2 plots illustrative forgetting functions showing the relation between log d and RI. The shape of such plots usually resembles some sort of decay function, and they are often modeled as hyperbolic or exponential decays. The hyperbolic decay function can be written

$$\log d_t = \log d_0 \frac{h}{h + t}, \qquad (4)$$

where t is the duration of the retention interval and $\log d_t$ is the accuracy at time t. The parameter $\log d_0$ estimates accuracy when $t = 0$, and the parameter h measures the rate of decay of $\log d_0$ (i.e., the half-life). The exponential decay function is similar, and it can be written

[1] A lengthy description of this model is unnecessary for the main purpose of this article. Interested readers should refer to Alsop and Davison (1991) or Davison and Nevin (1999).

$$\log d_t = \log d_0 \exp(-bt), \qquad (5)$$

where the parameter b measures the rate of exponential decay, and the remaining notation is as above. There is no clear consensus over which decay model better fits animals' performance; each fits some data sets better than the other, but often the variance accounted for by Equations 4 and 5 is very similar. In any case, Equations 4 and 5 provide equivalent conceptual frameworks. Two parameters summarize delayed matching-to-sample performance; the first parameter, log d_0, measures the overall difficulty of the discriminations at the 0-s RI, and the second parameter, h (Equation 4) or b (Equation 5), captures the memory component of the task (i.e., accuracy decreasing as a function of time).

Equations 4 and 5 are useful because they provide a relatively precise means to determine how an independent variable affects DMTS performance. For example, imagine an experiment in which animals performed a DMTS task with and without administration of some drug. Equations 4 and 5 allow the researcher to determine if the drug (a) changed the overall discrimination between the stimuli, but left the rate of forgetting unchanged; (b) left the overall discrimination unchanged, but altered the rate of forgetting; or (c) changed both the overall discrimination and the rate of forgetting. The effects of other independent variables (e.g., changing the intertrial interval, varying reinforcer magnitudes) on DMTS performance can be examined in the same way.

There is a problem with using Equations 3, 4, and 5 to plot and assess forgetting functions. Short-term memory studies usually arrange highly discriminable sample stimuli to avoid floor effects at longer RIs, especially if the researchers intend to compare forgetting functions following some change in procedure or the administration of a drug. As a result, subjects can be extremely accurate at the shorter RIs, and may make few or no errors. For example, Sargisson and White (2003) arranged a DMTS task in which the RIs were 0, 2, 4, 6, and 8 s, and the sample stimuli were a green key and red key. For each condition, they analyzed the data summed across the last five sessions, which gave a maximum of 80 trials at each RI. They replicated the baseline condition four times. The pigeons' accuracy was extremely high at the 0-s RI. Three of the 7 subjects never made an error in any of the four baseline sets of data (more than 300 trials in total), 1 subject made one error in total, and the remaining 3 subjects made three errors in total. Obviously, this level of accuracy demands a correction rule to conduct a signal-detection analysis; in this case the log-linear rule (K. G. White, personal communication).

To test the effect of correction rules on the analysis of DMTS data, Equation 4 was used to generate a series of true hyperbolic decay functions. The values chosen for log d_0 reflected the extremely high accuracy that subjects often show at short RIs. The maximum value of log d_0 was 3.0, which corresponds to an error rate of about one error every 1000 trials. Although this may seem an unrealistically high level of accuracy, empirical evidence indicates otherwise (e.g., Sargisson & White, 2003). Therefore, a log d_0 of 3.0 provides a reasonable upper limit. The two other values of log d_0 chosen for the analysis were less extreme, 2.0 (about one error every 100 trials) and 1.7 (about two errors every 100 trials). To provide a crude conversion to a more familiar metric, the three values of log d_0 cover a range of accuracy from approximately 98% correct (log d_0 = 1.7) to 99.9% correct (log d_0 = 3.0). The leftmost panel of Figure 2 plots the true values of log d_t at RIs of 0 s to 16 s for the three different values of log d_0, in each case assuming a half-life, h, of 5 s (Equation 2).

The next phase of the analysis involved simulating performance and comparing it to the true functions in Figure 2. The true functions provided expected ratios of correct to error responses at each RI for B_{11}/B_{12} and B_{22}/B_{21}. These expected ratios were converted to proportions, and then used in Monte Carlo simulations to generate numbers of responses for B_{11}, B_{12}, B_{21}, and B_{22} assuming 80 trials (40 of each sample stimulus) at each RI. This was a realistic number of trials at each RI from one condition of a DMTS experiment (e.g., Sargisson & White, 2003). These data were then adjusted using the two correction rules. In each case, estimates of log d (Equation 1) were then calculated at each RI. These simulations were run 100 times, and the mean estimates of log d across

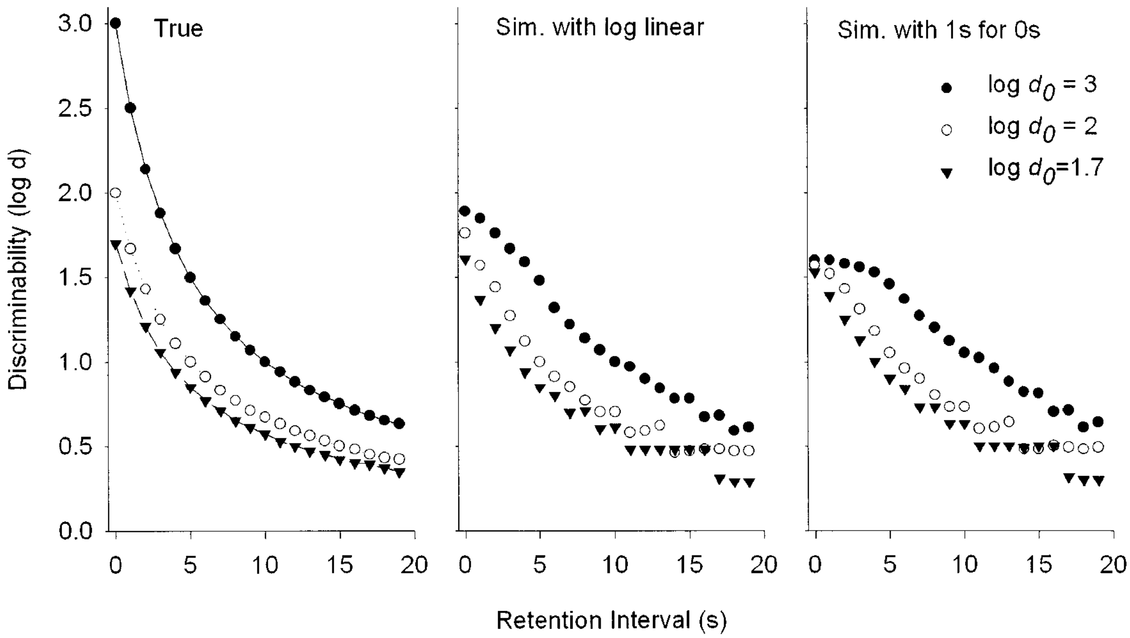

Fig. 2. The left panel shows true functions for Equation 4 when log d_0 was 3, 2, and 1.7, and the half-life, *h,* was 5 s. The middle and right panels show the means (and standard errors) following 100 simulations of these functions using the log-linear and 1s-for-0s rule, respectively.

these simulations are plotted as a function of RI in the middle (log-linear rule) and rightmost (1s-for-0s) panels of Figure 2.

Figure 2 shows that the forgetting functions obtained from the simulations with corrections for zero errors differed markedly from their respective true functions. In particular, the simulations consistently underestimated log d_t at shorter delays for the functions in which the true log d_0 was 3.0 and 2.0. It is clear that, in this example, the obtained forgetting functions provided a misleading representation of true performance, and that any conclusions based on the shape of these functions would require caution.

The differences between the true functions and the corrected simulations were also evident in quantitative analyses of the data in Figure 2. Table 1 shows the parameter estimates obtained after Equation 4 and Equation 5 were fitted to the results of the simulations using the Solver feature of MSExcel®. The true values (i.e., those used to generate the three functions in the leftmost panel of Figure 2) varied log d_0 and held the half-life, *h,* constant. The parameter estimates obtained from fits to the simulations painted a different picture. For the simulations with the log-linear rule applied, the fits suggested that the functions differed in terms of both log d_0

Table 1

The results of fitting hyperbolic and exponential decays (Equations 4 and 5) to the forgetting functions produced by the simulations using either the log linear rule or the 1s-for-0s rule.

True values		Hyperbolic fits (Equation 4)						Exponential fits (Equation 5)					
		Simulations with log linear			Simulations with 1s-for-0s			Simulations with log linear			Simulations with 1s-for-0s		
Log d_0	Half-life	Log d_0	Half-life	VAC	Log d_0	Half-life	VAC	Log d_0	b	VAC	Log d_0	b	VAC
3.00	5.0	2.04	10.0	0.96	1.77	14.9	0.91	1.97	0.07	0.99	1.74	0.05	0.96
2.00	5.0	1.83	5.9	0.99	1.71	7.3	0.97	1.69	0.09	0.98	1.62	0.08	0.99
1.70	5.0	1.61	5.7	0.99	1.58	6.4	0.99	1.47	0.09	0.96	1.47	0.09	0.97

and h. For the simulations using 1s-for-0s rule, the situation was more extreme. The major difference among the three functions was in terms of the half-life, h, whereas the estimates of log d_0 were relatively similar; that is, the simulations obtained the opposite relation to that used to generate the true functions.

Figure 2 and Table 1 pose problems for research that uses Equation 4 or Equation 5 to evaluate the effects of various interventions on animal short-term memory. Consider the earlier example of a study that investigated the effects of a drug on DMTS performance. Let us assume that, in reality, a certain drug had no effect on the rate of forgetting (h in Equation 4), but it caused a small decrease in overall accuracy on the task, log d_0, by lowering motivation or attention. In other words, the true effect of the drug is similar to the difference between the functions with log d_0 = 3.0 and log d_0 = 1.7 shown in the leftmost panel of Figure 2. The results of the present analysis, however, indicate that an actual experiment could obtain results similar to the simulations with the two correction rules (Figure 2, Table 1); that is, the researchers would find that the drug had a significant effect on the rate of forgetting, and little or no effect on overall accuracy—the opposite of the true effect. Notice that in this hypothetical example the problem arises more because the baseline condition has been incorrectly estimated rather than any difficulties with the drug condition.

The experiment by Sargisson and White (2003), mentioned earlier, provides a hypothetical example of what might occur when correction rules must be relied upon. Their baseline conditions arranged five RIs (0 s, 2 s, 4 s, 6 s, and 8 s), and reinforcers for correct responses were delivered immediately. They compared these baseline conditions with conditions in which reinforcers for correct responses were delayed. Figure 3 (left panel) shows the mean results from the baseline conditions and the conditions with the most extreme reinforcer delay (i.e., 8 s). Sargisson and White's analysis found that the rate of decay was greater at longer reinforcer delays than shorter reinforcer delays. This was an important empirical finding because it was inconsistent with the predictions of one model (i.e., Davison & Nevin, 1999), but it was consistent with a modified version of White and Wixted's (1999) competing model. This conclusion requires caution, however. As previously discussed, accuracy at the shortest delay during baseline was, for practical purposes, at ceiling, and this ceiling was determined largely by the number of trials (40 per stimulus) and the type of correction rule employed (e.g., Figures 1 and 2). To illustrate this point, Sargisson and White's data were reanalyzed but the values for B_{11}, B_{12}, B_{21}, and B_{22} were projected to 1000 trials per stimulus by multiplying each response value by 25. Although this projection altered the absolute numbers of responses, it preserved the obtained ratio of responses in each case. Figure 3 (right panel) shows the results of these projections. Although the estimates of discriminability at the 4-s, 6-s, and 8-s RIs were very similar across the two panels, the estimates at the shorter RIs were noticeably greater for the projected data, and more important, the difference between discriminability at the shorter RIs was about twice that obtained at the longer RIs; in other words, the rate of decay for the 0-s reinforcer conditions had increased and the two functions were more consistent with the predictions of Davison and Nevin. This is not to say that Sargisson and White's original conclusions were necessarily wrong. If they had extended actual training for as many trials as the projected data, their results might have been unchanged. The important issue is that the low number of trials at each RI and high overall accuracy in their study makes it difficult to establish any firm conclusions. Correction rules allowed some sort of analysis, but not a particularly convincing one.

It is unclear to what extent previous research might have reached questionable conclusions because of correction rules. Some studies do not provide sufficient data to determine whether there has been a problem with zero error rates or not. Other studies do not report how they dealt with the issue of zero errors. Even in cases in which information concerning correction rules is provided, there is still a problem; there is no way of knowing what the underlying true functions were in those studies (e.g., Figures 1 and 2). The conclusions of studies that used a correction rule could be correct, but they could as easily be completely wrong, and there is no obvious way to distinguish between the two.

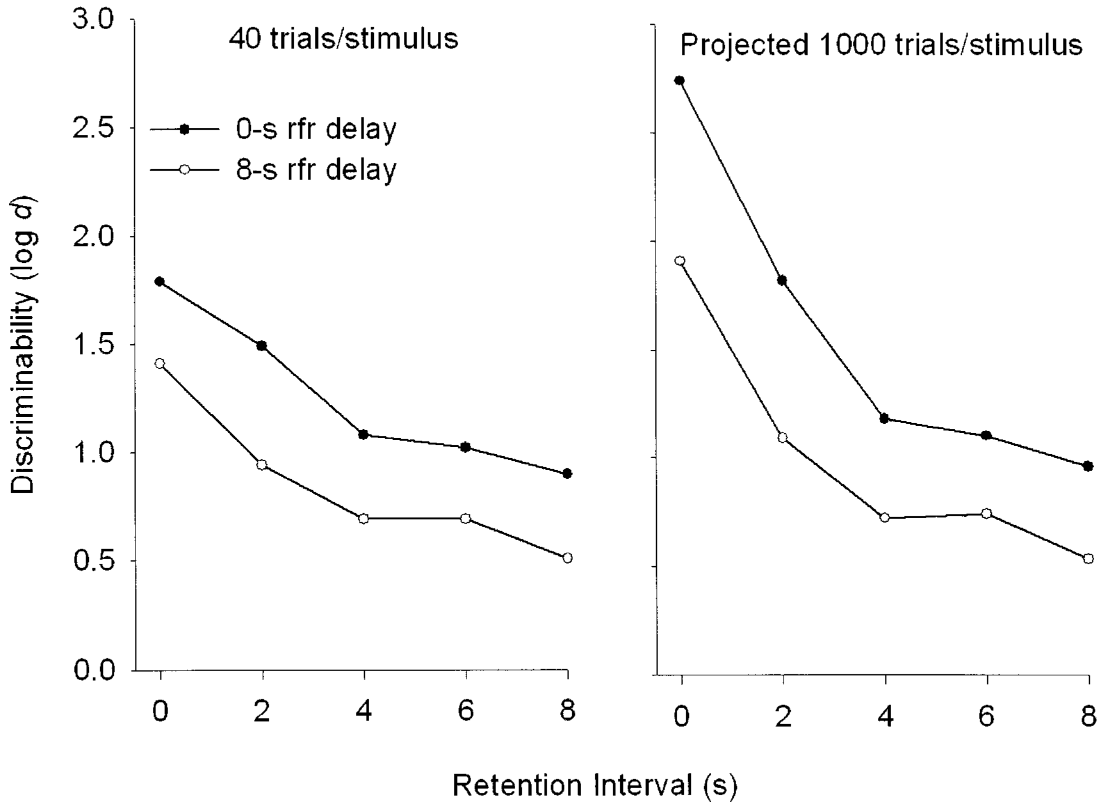

Fig. 3. The left panel plots mean discriminability as a function of retention interval separately from conditions with a 0-s reinforcer delay and an 8-s reinforcer delay (Sargisson & White, 2003). The right panel plots the same functions when the number of trials per stimulus was projected to 1000.

As a rule of thumb, any forgetting function should be treated with caution when the mean accuracy at the shortest RI is at, or close to, the maximum calculable accuracy given the number of trials. Under these circumstances, it is likely that all or some of the subjects' data required corrections, and that the shape of the function and any derived higher-order parameter estimates (Equations 4 and 5) could be misleading.

There is no easy solution for this analytical problem, but an important contributing factor is the number of trials on which calculations are based. For example, an experiment might arrange 80 to 100 trials each session and analyze response totals over the last five sessions. Although this sums to 400 to 500 trials, this total is a little misleading. If, as is often the case, there are several different RIs arranged within each session, the number of trials at each RI remains modest (e.g., Sargisson & White, 2003). Figure 4 plots the true function and the means following 100 sets of simulations of 80 trials at each RI using the log-linear rule (Figure 2) and adds plots of the means following 100 sets of simulations following 200 and 500 trials at each RI. Greater numbers of trials decreased the discrepancy between the true function and the simulations, and so reduced the chances of reaching an erroneous conclusion. It did not completely remove the risk, however.

The effect of the number of trials on the shape of forgetting functions (Figure 4) also might contribute to the debate concerning whether exponential or hyperbolic decay models better describe DMTS performance. Studies that prefer the hyperbolic decay function often have arranged only one retention interval per condition, and then have run four or five conditions to arrange enough retention intervals to fit a forgetting function

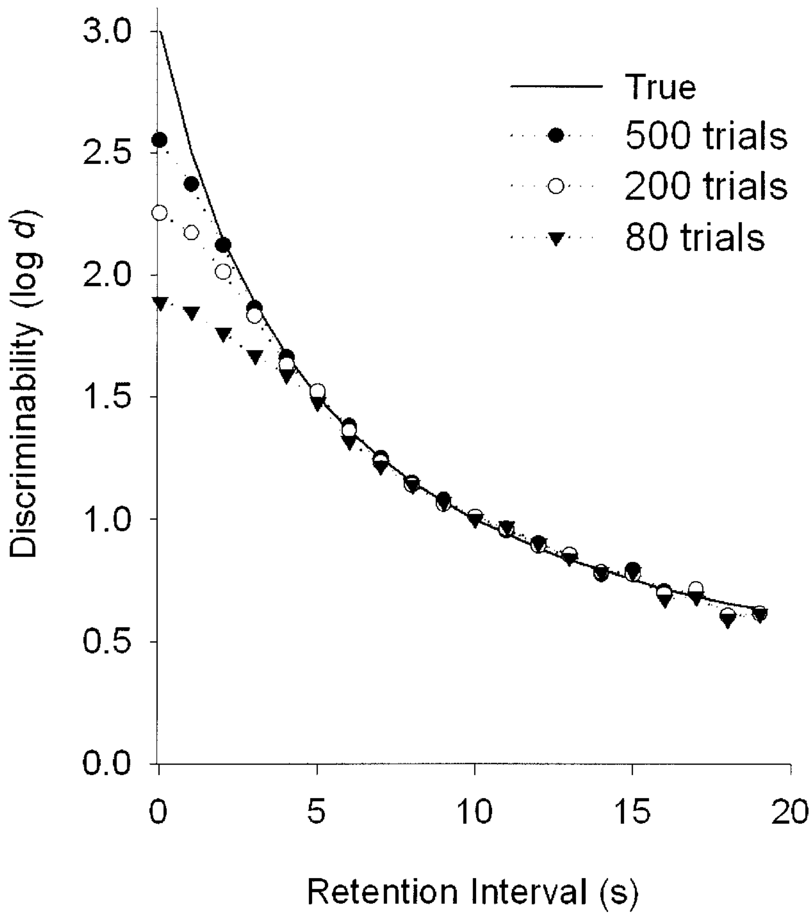

Fig. 4. Discriminability is plotted as a function of retention interval for the true function and for simulations using different numbers of trials per simulation.

(e.g., Harnett, McCarthy, & Davison, 1984). This means that there were several hundred trials per retention interval. Studies that prefer the exponential decay function have typically used a mixture of retention intervals in one condition (e.g., Sargisson & White, 2003). Although this allows a quick determination of a forgetting function, it reduces the number of trials per retention interval. Consider Figure 2, in which a hyperbolic decay function was used to generate the simulated functions. Despite this fact, Table 1 shows that the hyperbolic decay model (Equation 4) did not always provide the best fit following the simulations with 80 trials per retention interval; at higher discriminabilities, the exponential function (Equation 5) accounted for a greater proportion of the variance in the data.

Reinforcer Control in Detection and Delayed Matching-to-Sample Procedures

The previous section focused on the effects of correction rules on quantitative analyses of the discriminability between sample stimuli. Here I will consider the effect of correction rules on estimates of response bias, the tendency of a subject to favor one response over the other irrespective of the sample stimuli. A response bias can be produced in a conditional discrimination or a DMTS procedure by reinforcing one response alternative more frequently (or with larger reinforcers) than the other (e.g., McCarthy & Davison, 1979).

Algebraic combination of Equations 1 and 2 shows that the generalized matching law describes how response bias should change as

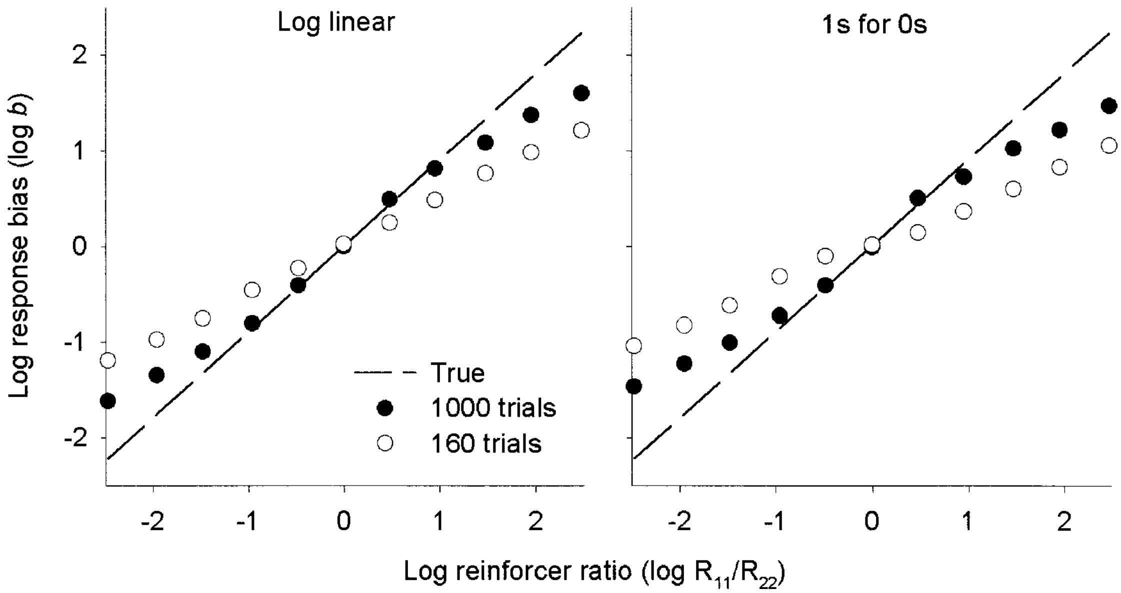

Fig. 5. Response bias is plotted as a function of the log reinforcer ratio for the true values (dashed line) and after the log-linear rule (left panel) and 1s-for-0s rule (right panel) were applied to simulated data (see text for details).

reinforcer ratios are varied across conditions (Davison & Tustin, 1978). This is usually presented in its logarithmic form,

$$\frac{1}{2}\log\left(\frac{B_{11}B_{21}}{B_{12}B_{22}}\right) = a\log\left(\frac{R_{11}}{R_{22}}\right) + \log c, \quad (6)$$

where all notation is as above. Equation 6 makes two important predictions. First, it predicts that, for a given level of stimulus discriminability, there is a linear relation between response bias and the reinforcer ratio. Second, it predicts that the relation between response bias and the reinforcer ratio should be independent of changes in stimulus discriminability (d in Equations 1 and 2).

To test the effect of correction rules on the linearity of Equation 6, Equations 1 and 2 were used to generate values of B_{11}, B_{12}, B_{21}, and B_{22} when d = 100, a = 0.9, and the reinforcer ratio varied from 300:1 to 1:300. This was done to produce a series of values for 20 simulations assuming 160 trials per condition, and a series of values assuming 800 trials per condition. The log-linear rule and the 1s-for-0s rule were applied to these values and to the results of the simulations in the manner described above.

Figure 5 plots the predicted relation (Equation 6) between response bias and the reinforcer ratio (dashed line), and the obtained relations following application of the log-linear rule (left panel) and the 1s-for-0s rule to the simulations (right panel). Following both correction rules, the simulated data show an orderly ogival deviation from the predicted line. Fewer trials per simulation led to greater deviations. Although the deviations were more pronounced at more extreme reinforcer ratios, the separation was apparent at reinforcer ratios commonly arranged in detection experiments (i.e., $-1 < \log(R_{11}/R_{12}) < 1$). Taken at face value, the simulations in Figure 5 would indicate that Equation 6 was wrong because the obtained relations were not linear. In fact, the ogives shown in Figure 5 are more consistent with predictions of Alsop and Davison's (1991) model of detection (for example, see Figure 6 of Davison & Nevin, 1999) than Equation 6.

To test the effect of the correction rules on the independence between sensitivity to the reinforcer ratio (a in Equation 6) and discriminability, Equations 1 and 2 were used to generate values of B_{11}, B_{12}, B_{21}, and B_{22} for five different reinforcer ratios (9:1, 3:1, 1:1, 1:3, 1:9) for each of thirteen different levels of stimulus discriminability (i.e., d = 1, 2, 4, 7, 10, 20, 40, 70, 100, 200, 400, 700, or 1000).

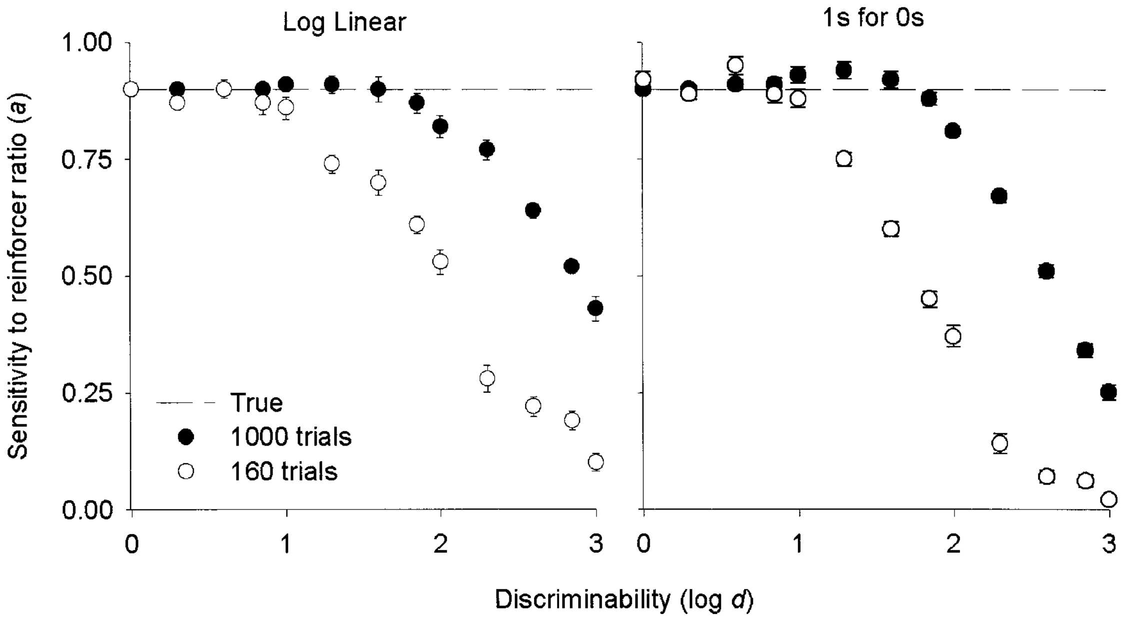

Fig. 6. Sensitivity to the reinforcer ratio is plotted as a function of the discriminability between the sample stimuli for the true values (dashed line) and after the log-linear rule (left panel) and 1s-for-0s rule (right panel) were applied to simulated data (see text for details).

At each level of discriminability, Equation 6 was fitted to the hypothetical data from the five different reinforcer ratios. These simulations were run 20 times assuming 1000 trials per condition, and assuming 160 trials per condition.

Figure 6 plots the mean sensitivity of behavior to changes in the reinforcer ratio (i.e., a in Equation 6) as a function of assigned stimulus discriminability (log d). The dashed line shows the true value of a, which is 0.9 across levels of discriminability by definition. The simulations, on the other hand, show an inverse relation between discriminability and sensitivity to the reinforcer ratio. For the simulations of 160 trials per condition, this effect was readily apparent at reasonably modest levels of discriminability (i.e., log $d > 1$). For the simulations of 1000 trials per condition, this covariation was most pronounced at higher levels of discriminability (i.e., log $d > 2$).

The results of the simulations shown in Figure 6 show a correlation between parameters that, in theory, should be independent (Davison & Tustin, 1978). Such a failure of parameter invariance usually indicates that the underlying model is incorrect (e.g., Nevin, 1984). Data from an actual experiment that looked like the simulations in Figure 6 probably would be interpreted as inconsistent with Davison and Tustin's behavioral model of signal detection, and more consistent with a model such as Alsop and Davison (1991).

Jones and White (1992) conducted an experiment that can be examined in light of the results shown in Figure 6. Their pigeons performed a DMTS task with four different RIs arranged in each condition. Across conditions, Jones and White varied the relative distribution of reinforcers for the two types of correct response. They used Equation 6 to analyze the effect of changes in the reinforcer ratio on response bias at each RI, and they used the 1s-for-0s rule for instances of zero errors. The relation between sensitivity to reinforcement, a, and discriminability, log d, is shown in Figure 7 (left panel). There was a significant decrease in a as log d increased, consistent with both the Alsop and Davison (1991) and White and Wixted (1999) models of signal detection. A similar result was obtained using the log-linear rule (Figure 7, middle panel). There were, however, only 80 trials per RI for each stimulus. To assess what effect this relatively low number of trials

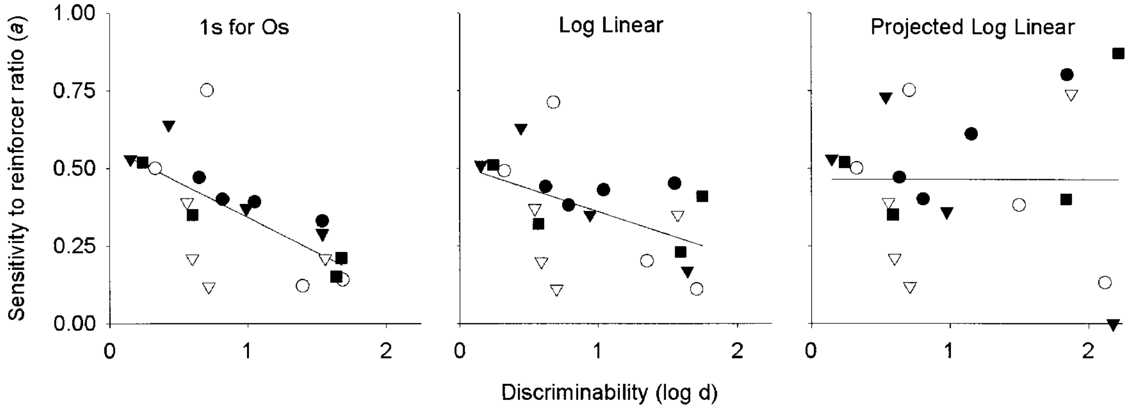

Fig. 7. The left panel plots sensitivity to the reinforcer ratio (a in Equation 6) as a function of discriminability following the 1s-for-0s rule for Jones and White's (1992) data. The middle panel shows the same data following the log-linear rule. The right panel shows the results when the number of trials per retention interval was projected to 1000. The different symbols represent the parameter estimates from individual subjects.

might have on the analysis, projected data for 1000 trials per RI for each stimulus were created in the same manner to those shown in Figure 3. These data were analyzed using the log-linear rule, and Figure 7 (right panel) shows the results. Now there is no clear evidence of an interaction between sensitivity to the reinforcer ratio and stimulus discriminability. This projection and analysis requires the usual caveat. If Jones and White had run more trials, they might still have obtained the interaction shown in the left panel of Figure 7; there is no definitive way to separate the pigeons' actual performance from the contribution of the correction rule.

Conclusions

Extreme conditions are useful for comparing models because these are the situations in which models often make clearly distinguishable predictions. Unfortunately, in the case of conditional discriminations and DMTS procedures, these conditions increase the likelihood of ceilings on performance (i.e., zero errors). If all that is needed is an ordinal comparison between two parameter estimates, then a technique such as the log-linear rule usually can be applied with some confidence. If, however, more sophisticated quantitative analyses are required, analyses for which absolute values are important (e.g., Equations 1 to 6), then correction rules can be false friends. In terms of the signal-detection analyses, the true functions generated by the Davison and Tustin (1978) model predicted an independence of discriminability with changes in the reinforcer distribution (Equation 3), a linear relation between response bias and the reinforcer ratio (Equation 6), and an independence between sensitivity to the reinforcer ratio and discriminability (Equation 6). The simulations found that discriminability covaried with changes in the reinforcer distribution (Figure 1), the relation between response bias and the reinforcer ratio was an ogive (Figure 5), and sensitivity to the reinforcer ratio was inversely related to discriminability (Figure 6). In terms of DMTS performance, hypothetical functions with identical rates of decay and different initial discriminabilities were transformed into functions with similar initial discriminabilities and different rates of decay (Figure 2, Table 1).

It is virtually impossible to identify those past studies whose results were definitely affected by the use of correction rules. First, the raw data are not always available, and the authors do not always state how they dealt with any problems concerning zero errors. Second, even if it was clear that a correction rule had been used, there is no clear way of determining how badly the results were affected. For example, the projected data in Figures 3 and 7 only show what *might* have been the case for Sargisson and White's (2003) and Jones and White's (1992) studies. The projected data do not provide a definitive reanalysis. That said, the simulations in

Figures 1, 2, 4, 5, and 6 indicate that results from past studies that arranged relatively few trials per data point and obtained high levels of accuracy probably require some caution; there is a real risk of drawing misleading conclusions.

How should behavioral models of signal detection (or related work with conditional discriminations) address the issue of zero errors at higher discriminabilities? There is no straightforward answer. Figures 1, 4, 5, and 6 clearly indicate that running a large number of trials per data point will attenuate the problem. If an experimenter wants to compare the predictions of models at extreme conditions, the simulations indicate that one or two thousand, rather than a few hundred, trials are necessary. Even under these conditions, a correction rule could introduce deviations in the data, but at least these effects will be smaller. Another strategy is to avoid conditions in which extreme performance is likely, unless such conditions are crucial for the experimental question. For example, a subject discriminating between stimuli with log d = 1.7 and showing no response bias would be correct on about 98% of the trials; in other words, it would only make about 20 errors every 1000 trials. At such low error rates, a small degree of variability in performance can produce quite large changes in derived parameters, and of course, the problem simply gets worse at higher discriminabilities.

Throughout the present article, a number of different models of detection (Alsop & Davison, 1991; Davison & Tustin, 1978; White & Wixted, 1999) and DMTS performance (Equations 4 and 5) have been mentioned. The present article is not intended to promote one of these models over the others. The Davison and Tustin (1978) model was used as the basis for generating true functions because it makes very clear predictions and it is probably the model with which most readers are most familiar. The finding that deviations in the simulated data were ordinally consistent with the Alsop and Davison model (e.g., Figures 1, 5, and 6) suggests that experiments trying to distinguish between this model and Davison and Tustin's model require particular care. Likewise, quantitative comparisons of DMTS performance and models need a degree of caution. The log-linear rule and the 1s-for-0s rule provide, at best, an educated guess about what data might look like, and need to be recognized as such.

REFERENCES

Alsop, B., & Davison, M. (1991). Effects of varying stimulus disparity and the reinforcer ratio in concurrent schedule and signal-detection procedures. *Journal of the Experimental Analysis of Behavior, 56,* 67–80.

Baum, W. M. (1974). On two types of deviation from the matching law: Bias and undermatching. *Journal of the Experimental Analysis of Behavior, 22,* 231–242.

Davison, M., & McCarthy, D. (1988). *The matching law: A research review.* Hillsdale, NJ: Erlbaum.

Davison, M., & Nevin, J. A. (1999). Stimuli, reinforcers, and behavior: An integration. *Journal of the Experimental Analysis of Behavior, 71,* 439–482.

Davison, M., & Tustin, R. D. (1978). The relation between the generalized matching law and signal detection theory. *Journal of the Experimental Analysis of Behavior, 29,* 331–336.

Godfrey, R., & Davison, M. (1998). Effects of varying sample- and choice-stimulus disparity on symbolic matching-to-sample performance. *Journal of the Experimental Analysis of Behavior, 69,* 311–326.

Godfrey, R., & Davison, M. (1999). The effects of number of sample stimuli and number of choices in a detection task on measures of discriminability. *Journal of the Experimental Analysis of Behavior, 72,* 33–55.

Green, D. M., & Swets J. A. (1966). *Signal detection theory and psychophysics.* New York: Wiley.

Harnett, P., McCarthy, D., & Davison, M. (1984). Delayed signal detection, differential reinforcement, and short-term memory in the pigeon. *Journal of the Experimental Analysis of Behavior, 42,* 87–111.

Hautus, M. J. (1995). Corrections for extreme proportions and their biasing effects on estimated values of *d'. Behavior Research Methods, Instruments, & Computers, 27,* 46–51.

Jones, B. M., & White, K. G. (1992). Sample-stimulus discriminability and sensitivity to reinforcement in delayed matching to sample. *Journal of the Experimental Analysis of Behavior, 58,* 159–172.

Macmillan, N. A., & Kaplan, H. L. (1985). Detection theory analysis of group data: Estimating sensitivity from average hit and false-alarm rates. *Psychological Bulletin, 98,* 185–199.

McCarthy, D. C., & Davison, M. (1979). Signal probability, reinforcement, and signal detection. *Journal of the Experimental Analysis of Behavior, 32,* 373–386.

McCarthy, D., & Davison, M. (1980). Independence of sensitivity to relative reinforcement rate and discriminability in signal detection. *Journal of the Experimental Analysis of Behavior, 34,* 273–284.

Nevin, J. A. (1984). Quantitative analysis. *Journal of the Experimental Analysis of Behavior, 42,* 421–434.

Nevin, J. A., Milo, J., Odum, A. L., & Shahan, T. A. (2003). Accuracy of discrimination, rate of responding, and resistance to change. *Journal of the Experimental Analysis of Behavior, 79,* 307–321.

Sargisson, R. J., & White, K. G. (2003). The effects of reinforcer delays on the form of the forgetting function. *Journal of the Experimental Analysis of Behavior, 80,* 77–94.

Watson, J. E., & Blampied, N. M. (1989). Quantification of the effects of chlorpromazine on performance under delayed matching to sample in pigeons. *Journal of the Experimental Analysis of Behavior, 51,* 317–328.

White, K. G., & Wixted, J. T. (1999). Psychophysics of remembering. *Journal of the Experimental Analysis of Behavior, 71,* 91–113.

Received October 17, 2003
Final acceptance June 8, 2004